Supplementary Note

PMD Hypomethylation in immortalized cell lines

We saw PMD hypomethylation in almost all cultured cell lines except for ESCs, iPSCs and their derived cell lines (**Fig. 4** Group ESC). Interesting observations included: 1) hESCs (including H1, H9 and HUES64 and 4star) and most hESC-derived progenitor cells were heavily methylated without visually detectable PMD, most likely due to hyperactivity of DNMT3B[1,2]. The stark contrast between the primary ICM sample and the heavily methylated hESCs suggests that cultured hESCs may reflect a later stage of post-implantation embryonic development, where expression of the DNMT3A and DNMT3B methyltransferases can help to maintain high levels of DNA methylation despite prolonged culture (**Fig. 5a**). 2) Two H1-derived Mesenchymal Stem Cells (MSCs) showed clear PMD structure (**Supplementary Fig. 7a**). 3) iPSCs, also with active DNMT3B[3] and with very little loss of PMD methylation in most samples, had residual trace PMDs in some samples (e.g., the 19.11 cell line) with respect to fore-skin fibroblasts from which they originated (**Supplementary Fig. 7a**).

We also note that although both ESCs and the proliferative tumors were high in the expression of DNMT3s compared to other normal tissues of non-embryonic origin, the level of expression in ESCs was higher than the most proliferative tumors. For example, the expression of DNMT3B in H1 hESC was higher than other cancer cell lines and primary tissues assayed in the ENCODE project by over ten folds (**Supplementary Fig. 18a**). Embryonic Carcinoma, sharing a similar early embryonic origin with ESCs, also

had the highest expression of both DNMT3A and DNMT3B compared to other cancer types in TCGA (**Supplementary Fig. 18b**). Like hESCs, these embryonic carcinomas did not manifest strong PMD structures either (**Supplementary Fig. 12**). Since DNMTs are part of a large DNA replication program, the high DNMT3s in most proliferative tumors are passively driven by the fast cell turn-over of the cancer cells, while ESCs actively express DNMT3s to maintaining their pluripotency. This explains the seemingly contradictory observations of a strong PMD structure in the proliferative tumors and lack of PMD structure in ESCs, despite both having high DNMT3s. This is supported by the the high expression of other replication program component genes (such as UHRF1 and other cell cycle dependent genes) in the highly proliferating tumors with severe PMD hypomethylation (**Fig. 7g**).

Improved analysis of HMD/PMD structure

Our focus here has been on cell-type invariant PMDs, which were useful for investigating general properties of methylation loss over time. The 49% of the genome we identified as occurring within “Common PMDs” (using the SD>0.15 method) contains essentially all of the cell-type-invariant PMD regions that we identified previously[8]. We defined these PMDs by exploiting the inherent variance in PMD hypomethylation levels across large cohorts of samples, which was the only cross-sample feature bimodally distributed between HMDs and PMDs. Under this definition, for example, the core tumor group (containing only solid tumors) had almost the same degree of shared PMDs with blood malignancies (82%) as it did with other solid tumors not from the core set (85%) (**Supplementary Fig. 8**). We do note that the power of this method might not apply to

sample cohorts with little variation in hypomethylation levels, but it worked well for all the sample groups we examined here.

Our focus on common PMDs does not discount the importance of cell-type-specific PMDs. The work of our group and others showed that about 25% of PMDs were cell-type specific[4,5], and our results here do not conflict with that. Others have established that cell-type specific cancer PMDs can be associated with gene expression differences, and distinguish different molecular subtypes of medulloblastoma and Atypical Teratoid/Rhabdoid tumors[5–7]. Work from Fortin and Hansen showed that these cell-type-specific PMD differences corresponded to cell-type-specific topological domain and chromatin structure differences using Hi-C and DNase data from the same cell lines[8].

We observed deep PMD hypomethylation in the methylome of T cells from a 103-year-old individual (**Fig. 6a**). Interestingly, in a previous study the hypomethylation patterns could not be conclusively called as PMDs even for the 103 year old sample, likely due to the noise introduced by CpGs other than solo-WCGWs[10]. We expect that incorporation of solo-WCGW sequence features can be used to improve current methods for such cell-type-specific PMD detection, including kernel-based[11], HMM-based[12] and multi-scale based[13], and methods for methylation array data[8]. Explicitly modeling and subtracting PMD-related hypomethylation will reduce noise and enhance our ability to detect changes in TET-mediated demethylation processes affecting short-range elements such as promoters, enhancers, and insulators.

While the discovery of solo-WCGW CpGs was a significant advance, the ability to detect differential PMDs in normal cell types with low levels of methylation loss, will remain a challenge. This is an important challenge to tackle, as it may allow the identification of PMD-associated cell-of-origin markers in cancer, which can be combined with mutational-signature-based cell-of-origin markers[9]. PMD domain structure can also act as a useful proxy for 3D topological changes and other chromatin features in clinical disease samples where Hi-C or other direct mapping methods are not feasible due to the quantity or quality of intact chromatin available. PMDs also mark regions of gene silencing, and thus can help to infer the the gene expression history of the cells being sampled. For instance, Hovestadt *et al.* showed that PMDs in medulloblastoma tumors reflected subtype-specific expression silencing in normal brain precursor cells[14].

Stability of rank-based correlation between methylomes

We performed a rank-based analysis of 792 genomic 100kb bins from chromosome 16 (**Fig. 5**) to measure the HMD/PMD structure in normal tissues at different developmental stages. The rank correlations had only minor variations between replica or closely related samples (**Supplementary Fig. 19a**) and the patterns were stable when using bins from different chromosomes (**Supplementary Fig. 19b**)

Alternative explanation of PMD hypomethylation

While our analysis implicated replication timing as the most strongly associated genomic determinant of PMD methylation loss, replication timing is in practice very tightly linked to the Hi-C compartment "B" and the nuclear lamina based on our work and the work of others[14,15,16]. While the *re-methylation window* model is mechanistically attractive, we cannot rule out an alternative *nuclear localization model* (**Fig. 8g**), where methylation loss is due to compositional differences between the two nuclear compartments *independent of replication timing*, including differential activity of DNMTs or other chromatin regulatory factors. Indeed, various proteins are known to be regulated at the level of sub-nuclear compartment localization, such as TRIM28 (KAP-1)[17]. It should be noted that the link between DNMT3B and H3K36me3 has been primarily described in mouse ES cells, which express a different isoform of Dnmt3b. Therefore, it remains possible that other DNMTs also contribute to the high methylation levels within early replicating regions. DNMT3A would be such a candidate, given that early replicating regions become hypomethylated upon Dnmt3a loss in a mouse lung cancer model[18]. Recent work suggests that the heterochromatin and euchromatin nuclear compartments have a physical barrier created by liquid heterochromatin droplets formed by HP1-mediated phase separation[19,20].

Relevance of the PMD sequence signature to somatic and germline mutational landscape

To investigate any potential impact of the PMD sequence signature on introducing cytosine deamination mutations in the CpG dinucleotides, we studied the relative proportion of somatic mutations that are within certain tetranucleotide sequence

contexts and certain numbers of neighboring CpGs. We compared somatic CpG to TpG mutations reported in an early gastric cancer whole-genome sequencing experiment and indeed confirmed that solo-WCGWs within late replicating PMDs had a lower CpG to TpG mutation rate compared with other sequence context (**Supplementary Fig. 16a**). However, we also observed higher somatic mutation density overall in PMDs compared to HMDs, confirming earlier reports[21], possibly due to compensating effect from transcription-coupled DNA repair[22]. More systematic investigation incorporating differential repair efficiencies will be necessary to investigate the effects solo-WCGW hypomethylation may have in shaping the single nucleotide mutational signatures observed in cancer and in evolution.

While only a limited number of samples were available for gametogenesis, we observed dramatic PMD hypomethylation in at least one germline cell type, the Germinal Vesicle, M-I Oocyte (**Fig. 5b**). This opens the possibility that local sequence determinants, HMD/PMD structure, or H3K36me3 distribution may play a role in methylation-sensitive deamination rates in the germline, and thereby help shape genome evolution. We studied *de novo* CpG->TpG mutations reported in a study of 1,548 Icelandic trios. We found that these *de novo* CpG->TpG mutations in the maternal germline were indeed depleted at CpGs in the WCGW context and with low local CpG density (**Supplementary Fig. 16b**). The trend is not as apparent in paternal *de novo* mutations, consistent with lack of strong PMD structure in sperm (**Fig. 5b**). The standing distribution of human and mouse CpGs is also consistent with the hypothesis that tendency of losing methylation in solo-WCGW context in the germline may exert a

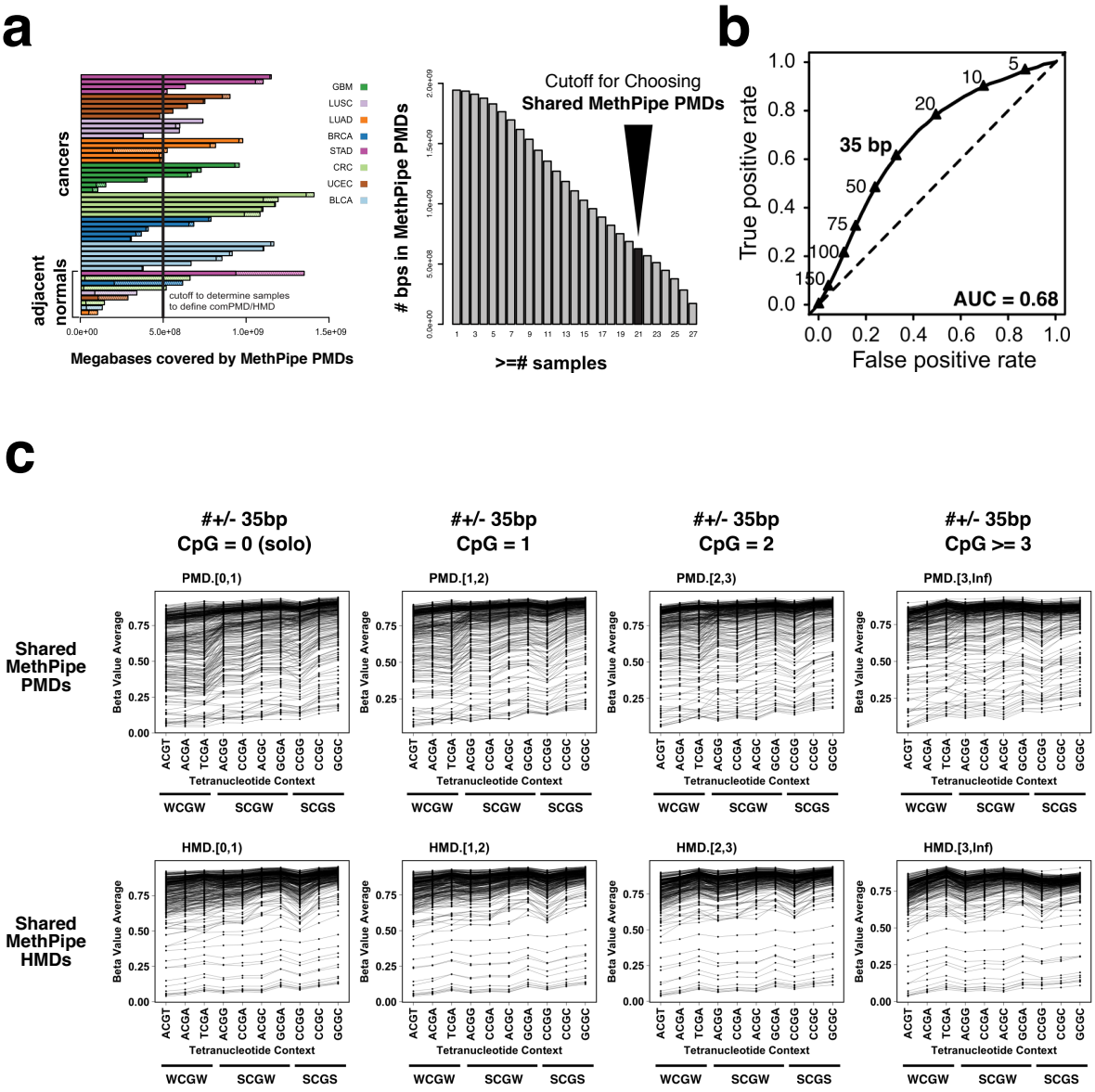
protective role for these CpGs against deamination (**Supplementary Fig. 16c-d**). Such mechanisms have been proposed for other mutational processes[23], and the well-defined genomic constraints on the hypomethylation process described here will allow these types of analysis.

Supplementary Note Reference

1. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99,** 247–257 (1999).
2. Laurent, L. *et al.* Dynamic changes in the human methylome during differentiation. *Genome Res.* **20,** 320–31 (2010).
3. Pawlak, M. & Jaenisch, R. De novo DNA methylation by Dnmt3a and Dnmt3b is dispensable for nuclear reprogramming of somatic cells to a pluripotent state. *Genes Dev.* **25,** 1035–1040 (2011).
4. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462,** 315–322 (2009).
5. Berman, B. P. *et al.* Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat. Genet.* **44,** 40–46 (2012).
6. Hovestadt, V. *et al.* Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* **510,** 537–541 (2014).
7. Johann, P. D. *et al.* Atypical Teratoid/Rhabdoid Tumors Are Comprised of Three Epigenetic Subgroups with Distinct Enhancer Landscapes. *Cancer Cell* **29,** 379–393 (2016).
8. Fortin, J.-P. & Hansen, K. D. Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. *Genome Biol.* **16,** 180 (2015).
9. Polak, P. *et al.* Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature* **518,** 360–364 (2015).
10. Vandiver, A. R. *et al.* Age and sun exposure-related widespread genomic blocks of hypomethylation in nonmalignant skin. *Genome Biol.* **16,** 80 (2015).
11. Hansen, K. D., Langmead, B. & Irizarry, R. a. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol.* **13,** R83 (2012).
12. Song, Q. *et al.* A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PLoS One* **8,** e81148 (2013).

13. Knijnenburg, T. a *et al.* Multiscale representation of genomic signals. *Nat. Methods* **11,** 689–94 (2014).
14. Shipony, Z. *et al.* Dynamic and static maintenance of epigenetic memory in pluripotent and somatic cells. *Nature* **513,** 115–119 (2014).
15. Hansen, R. S. *et al.* Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 139–44 (2010).
16. Pope, B. D. *et al.* Topologically associating domains are stable units of replication-timing regulation. *Nature* **515,** 402–405 (2014).
17. Iyengar, S. & Farnham, P. J. KAP1 protein: An enigmatic master regulator of the genome. *J. Biol. Chem.* **286,** 26267–26276 (2011).
18. Raddatz, G., Gao, Q., Bender, S., Jaenisch, R. & Lyko, F. Dnmt3a Protects Active Chromosome Domains against Cancer-Associated Hypomethylation. *PLoS Genet.* **8,** e1003146 (2012).
19. Strom, A. R. *et al.* Phase separation drives heterochromatin domain formation. *Nature* **547,** 241–245 (2017).
20. Larson, A. G. *et al.* Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* **547,** 236–240 (2017).
21. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499,** 214–8 (2013).
22. Hanawalt, P. C. & Spivak, G. Transcription-coupled DNA repair: two decades of progress and surprises. *Nat. Rev. Mol. Cell Biol.* **9,** 958–70 (2008).
23. Kenigsberg, E. *et al.* The mutation spectrum in genomic late replication domains shapes mammalian GC content. *Nucleic Acids Res.* **44,** 4222–4232 (2016).
24. Wang, K. *et al.* Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat. Genet.* **46,** 573–582 (2014).
25. Jónsson, H. *et al.* Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature* **549,** 519–522 (2017).

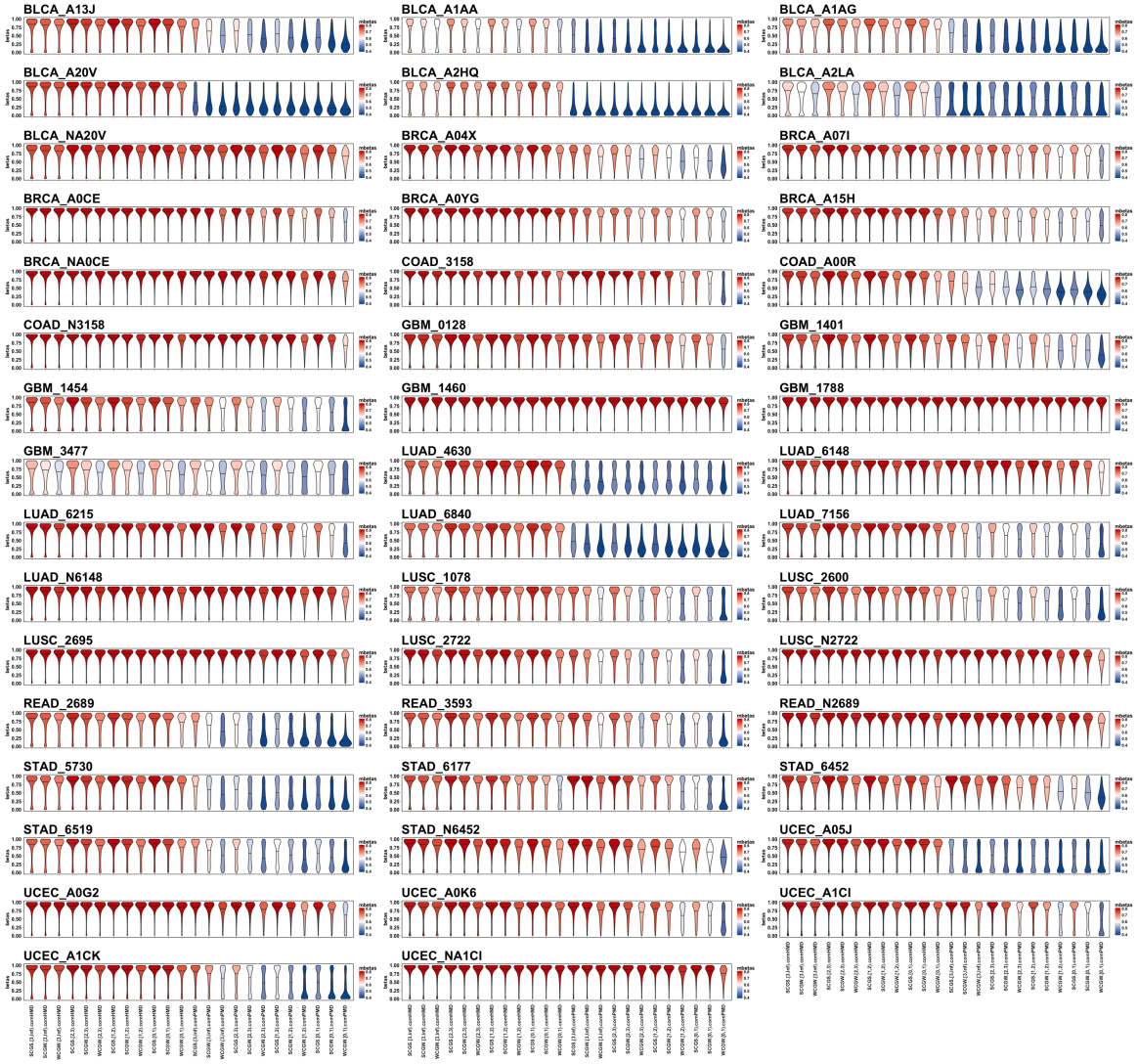
Supplementary Figures



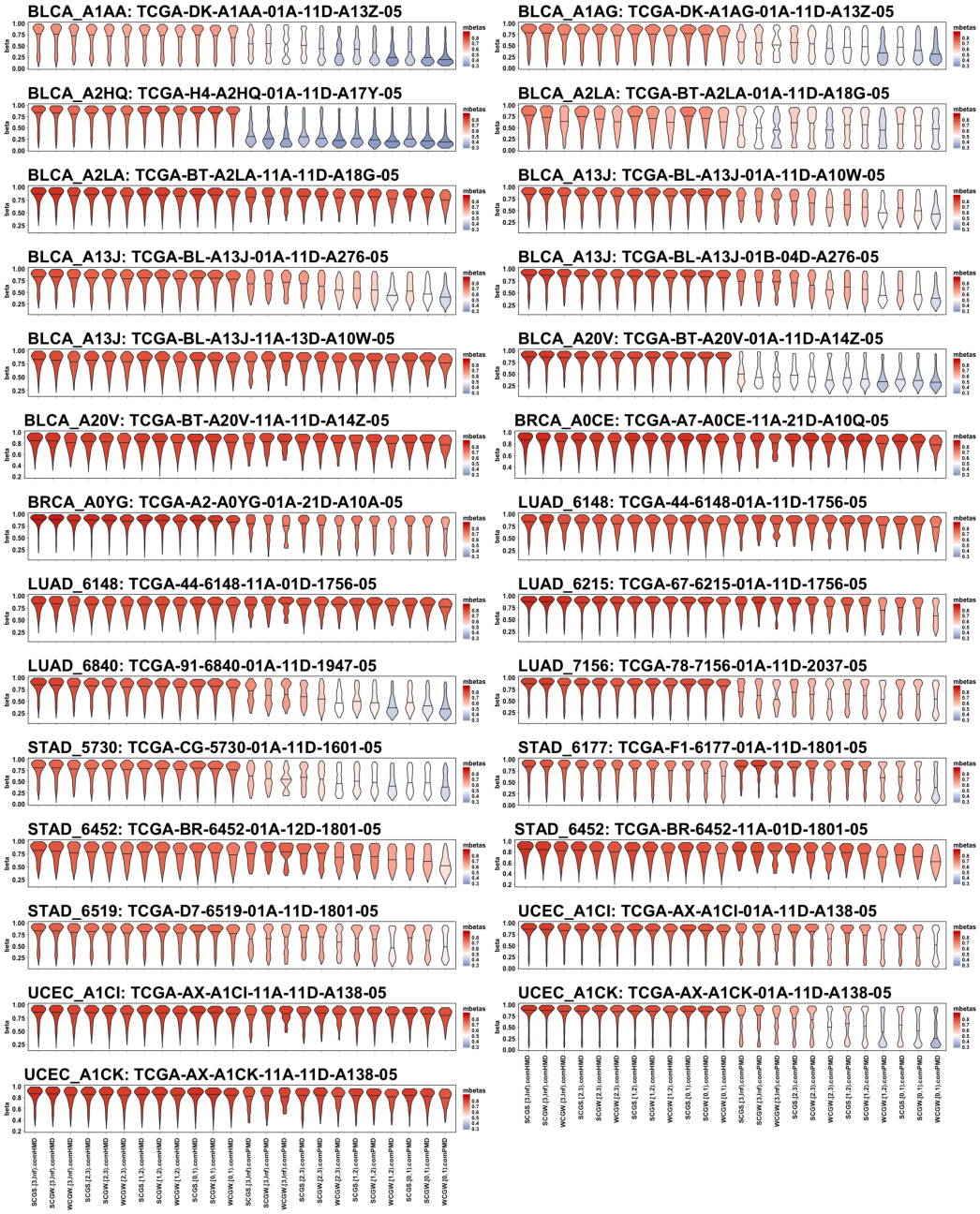
Supplementary Figure 1 (**a**) PMD calls by methpipe on tumor and adjacent normal samples reported in this study (left) and cutoff for choosing shared MethPipe PMDs (Note that this only used here and in Fig. 1, the definition of PMDs were updated later based on cross tumor SDs) from these methpipe calls (right). (**b**) Receiver Operating

Characteristic (ROC) curve showing prediction power of hypomethylation tendency with different sizes of the sequence window in defining Solo-CpGs in human (N=26,752,698 CpGs). (**c**) Methylation average of CpG dinucleotides in 10 tetranucleotide sequence context stratified by neighboring CpG number and genomic territory (PMD or HMD). Each panel includes 390 WGBS samples.

a

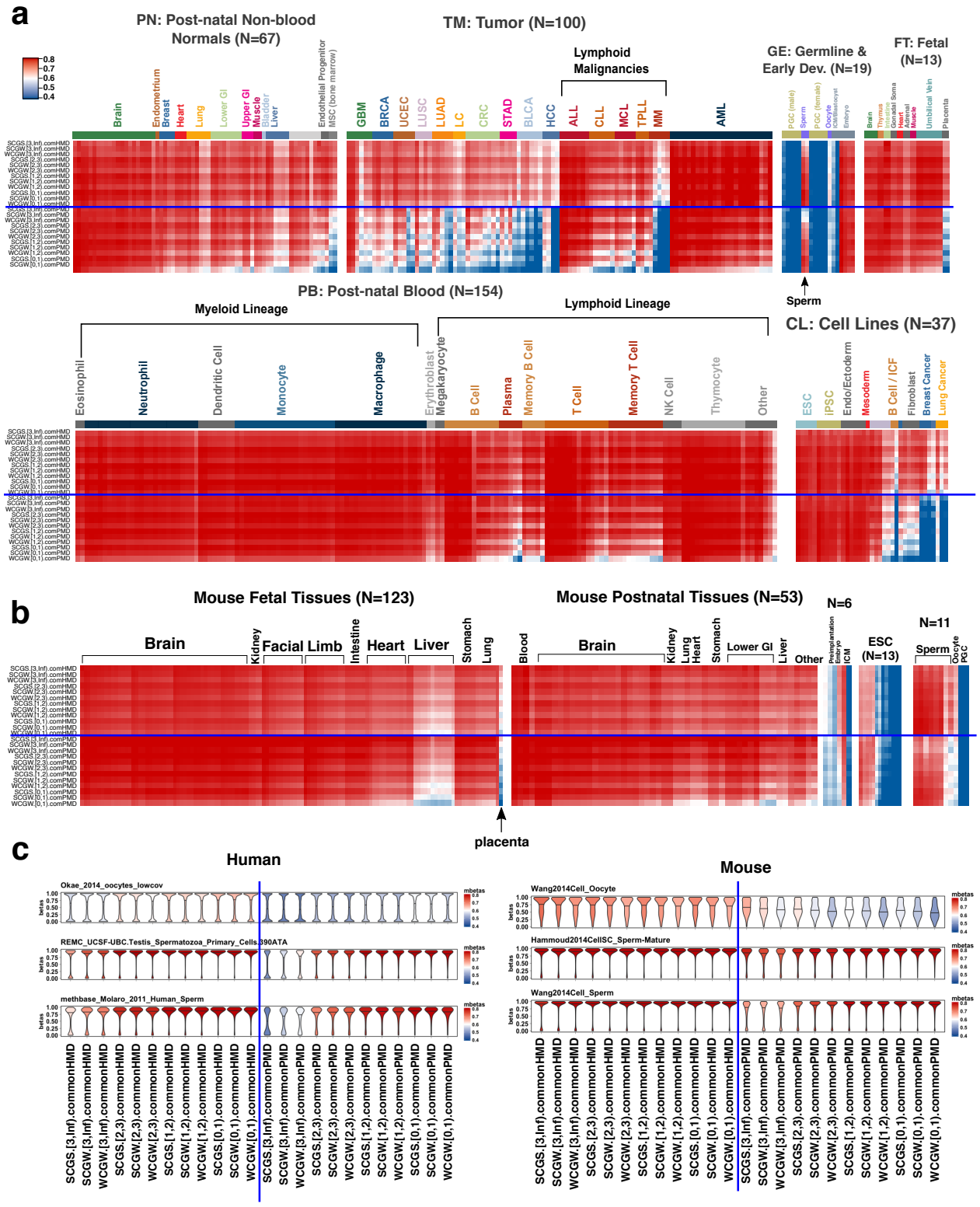

BLCA_A13J
BLCA_A1AA
BLCA_A1AG
BLCA_A20V
BLCA_A2HQ
BLCA_A2LA
BLCA_NA20V
BRCA_A04X
BRCA_A07I
BRCA_A0CE
BRCA_A0YG
BRCA_A15H
BRCA_NA0CE
COAD_3158
COAD_A00R
COAD_N3158
GBM_0128
GBM_1401
GBM_1454
GBM_1460
GBM_1788
GBM_3477
LUAD_4630
LUAD_6148
LUAD_6215
LUAD_6840
LUAD_7156
LUAD_N6148
LUSC_1078
LUSC_2600
LUSC_2695
LUSC_2722
LUSC_N2722
READ_2689
READ_3593
READ_N2689
STAD_5730
STAD_6177
STAD_6452
STAD_6519
STAD_N6452
UCEC_A05J
UCEC_A0G2
UCEC_A0K6
UCEC_A1CI
UCEC_A1CK
UCEC_NA1CI

b



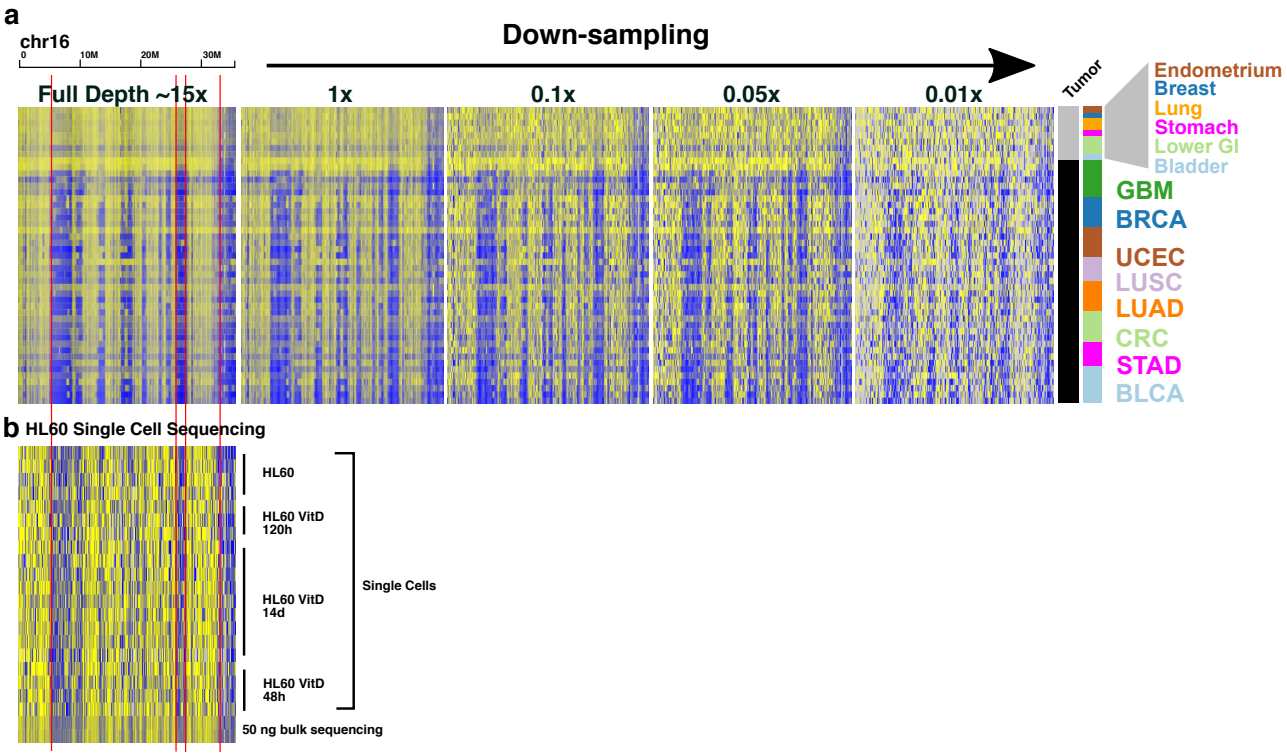
Supplementary Figure 2 (**a**) Violin plots of CpG methylation in 24 sequence contexts for all 47 TCGA WGBS samples (39 tumors and 8 normals) reported in this study. Elements of the violin plots represents the DNA methylation beta value of each CpG; (**b**) Methylation distribution of CpGs in 24 sequence contexts from 27 matched HM450 data

228 of the TCGA WGBS samples. Element of the violin plots represents the DNA
229 methylation beta value of each CpG.
230



Supplementary Figure 3 (**a**) Methylation average of CpG dinucleotides in 24 sequence contexts (rows) of 390 WGBS samples; (**b**) Methylation average of CpG dinucleotides in

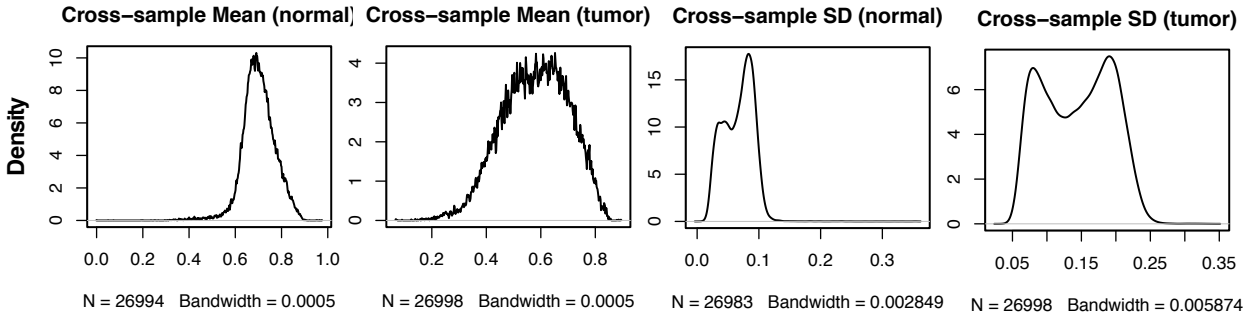
24 sequence context (rows) of 206 mouse WGBS samples; (**c**) Methylation distribution of CpG dinucleotides in 24 sequence contexts in one oocyte and two spermatozoa samples in human and in mouse respectively. N=26,752,698 CpGs for human and N=20,383,610 CpGs for mouse. Element of the violin plots represents the DNA methylation beta value of each CpG in the specific sequence context.



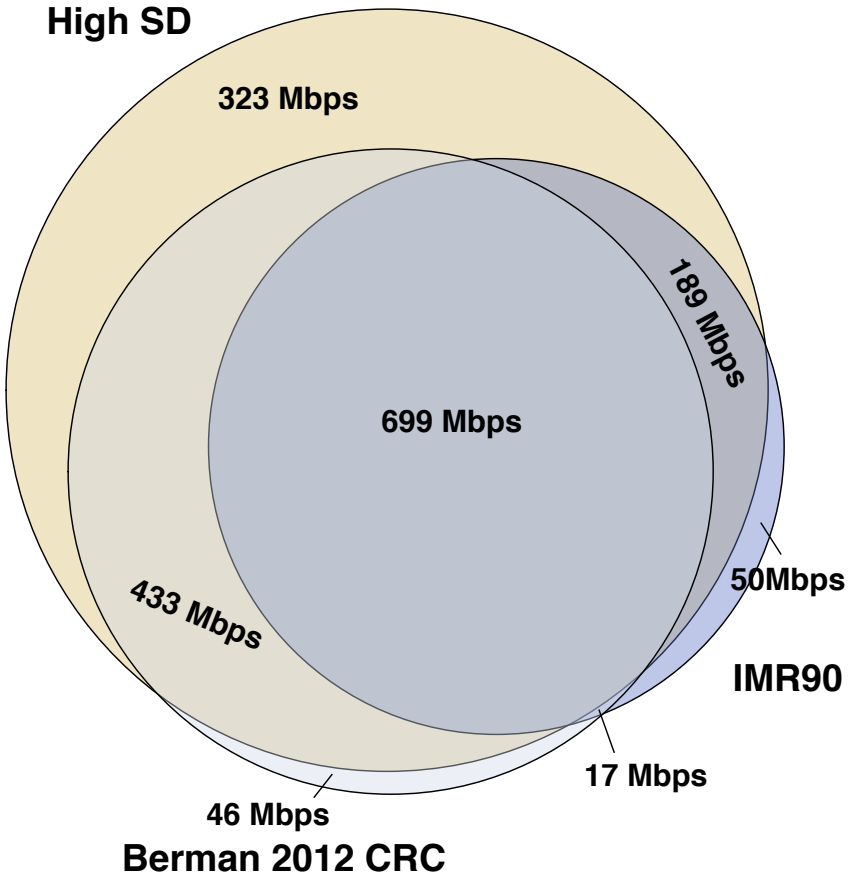
240

241 **Supplementary Figure 4** (**a**) Heatmap showing DNA methylation beta value of
242 chromosome 16p in 49 TCGA WGBS samples (40 tumors and 9 adjacent normal
243 samples, including colorectal cancer and matched normal from Berman et al. 2012
244 *Nature Genetics*) downsampled from 1x to 0.01x; (**b**) Heatmap showing DNA
245 methylation beta value of chromosome 16p in 20 single-cell whole genome bisulfite
246 sequencing (scWGBS) of HL60 cell line under vitamin D treatment as well as two bulk
247 WGBS data sets of 50ng (data from Farlik et al. 2015 *Cell Reports*, see also
248 **Supplementary Table 1**).

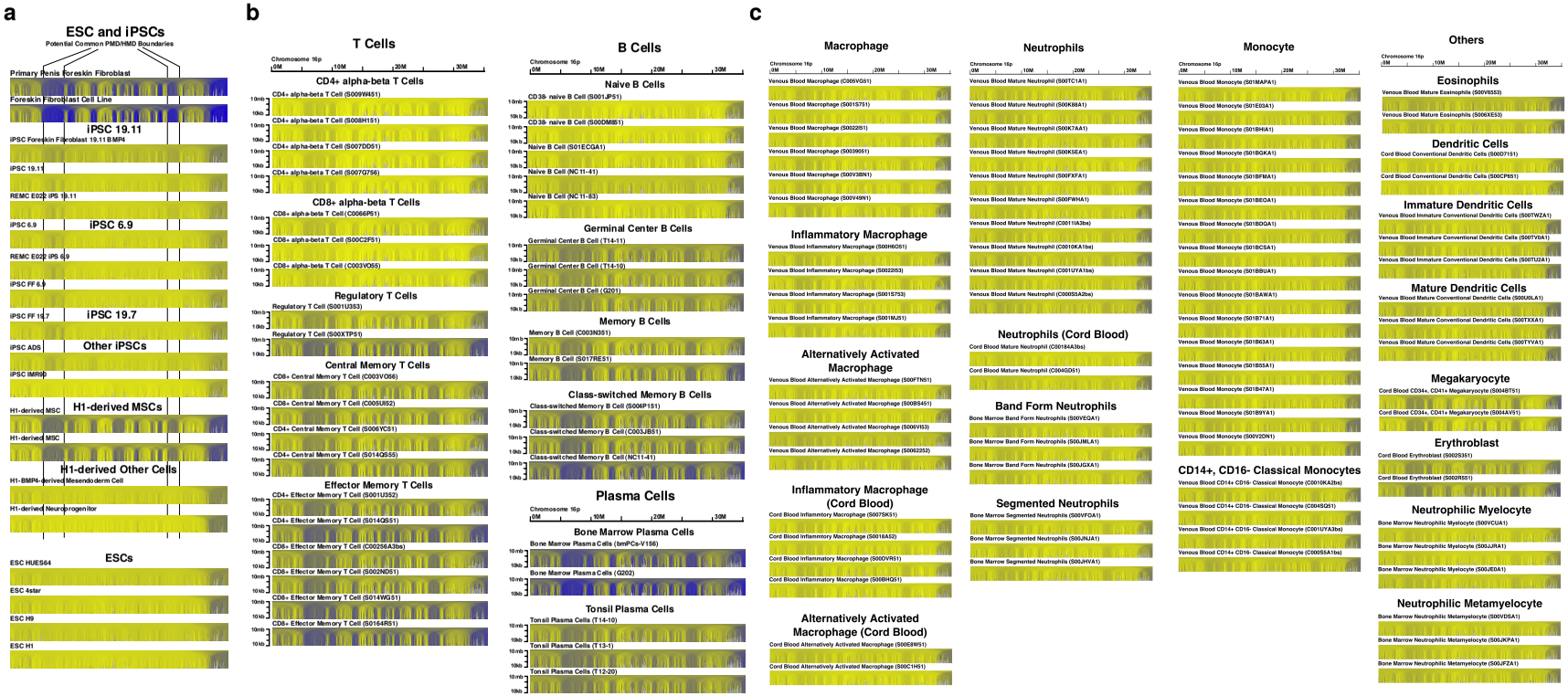
249



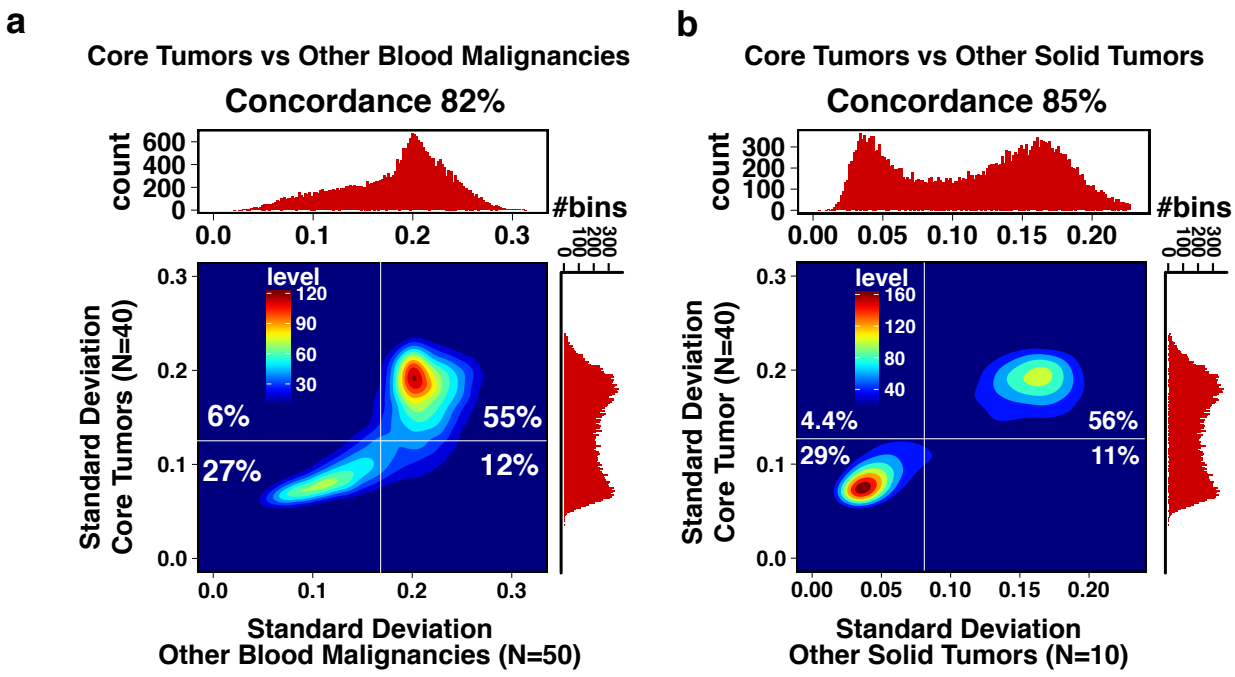
Supplementary Figure 5 Absence of bimodal distribution of cross-sample mean for the core normal and tumor WGBS samples;



Supplementary Figure 6 Overlap of PMD definition in this work with previous studies from colorectal cancer and IMR90 cell lines with overlapping area approximating numbers of overlapping base pairs.


a
ESC and iPSCs
Potential Common PMD/HMD Boundaries
Primary Penis Foreskin Fibroblast
Foreskin Fibroblast Cell Line
iPSC 19.11
IPSC Foreskin Fibroblast 19.11 BMP4
IPSC 19.11
REMC E022 IPS 19.11
IPSC 6.9
iPSC 6.9
REMC E022 IPS 6.9
IPSC FF 6.9
IPSC FF 19.7
iPSC 19.7
IPSC ADS
Other iPSCs
IPSC IMR90
H1-derived MSC
H1-derived MSCs
H1-derived MSC
H1-derived Other Cells
H1-BMP4-derived Mesendoderm Cell
H1-derived Neuroprogenitor
ESCs
ESC HUES64
ESC 4star
ESC H9
ESC H1
b
T Cells
Chromosome 16p
0M
10M
20M
30M
CD4+ alpha-beta T Cells
CD4+ alpha-beta T Cell (S009W451)
CD4+ alpha-beta T Cell (S008H151)
CD4+ alpha-beta T Cell (S007DD51)
CD4+ alpha-beta T Cell (S007Q756)
CD8+ alpha-beta T Cells
CD8+ alpha-beta T Cell (C0066P51)
CD8+ alpha-beta T Cell (S00C2F51)
CD8+ alpha-beta T Cell (C003VO55)
Regulatory T Cells
Regulatory T Cell (S001U353)
Regulatory T Cell (S00XTP51)
Central Memory T Cells
CD8+ Central Memory T Cell (C003VO66)
CD8+ Central Memory T Cell (C005UI52)
CD4+ Central Memory T Cell (S006YC51)
CD4+ Central Memory T Cell (S014QS55)
Effector Memory T Cells
CD4+ Effector Memory T Cell (S001U352)
CD4+ Effector Memory T Cell (S014QS51)
CD8+ Effector Memory T Cell (C00256A3bs)
CD8+ Effector Memory T Cell (S002ND51)
CD8+ Effector Memory T Cell (S014WG51)
CD8+ Effector Memory T Cell (S0164R51)
B Cells
Naive B Cells
CD38- naive B Cell (S001JP51)
CD38- naive B Cell (S00DM851)
Naive B Cell (S01ECGA1)
Naive B Cell (NC11-41)
Naive B Cell (NC11-83)
Germinal Center B Cells
Germinal Center B Cell (T14-11)
Germinal Center B Cell (T14-10)
Germinal Center B Cell (Q201)
Memory B Cells
Memory B Cell (C003N351)
Memory B Cell (S017RE51)
Class-switched Memory B Cells
Class-switched Memory B Cell (S004P151)
Class-switched Memory B Cell (C003JB51)
Class-switched Memory B Cell (NC11-41)
Plasma Cells
Bone Marrow Plasma Cells
Bone Marrow Plasma Cells (bmPCs-V156)
Bone Marrow Plasma Cells (Q202)
Tonsil Plasma Cells
Tonsil Plasma Cells (T14-10)
Tonsil Plasma Cells (T13-1)
Tonsil Plasma Cells (T12-20)
c
Macrophage
Venous Blood Macrophage (C005VG51)
Venous Blood Macrophage (S001S751)
Venous Blood Macrophage (S0022251)
Venous Blood Macrophage (S0039351)
Venous Blood Macrophage (S00V3BN1)
Venous Blood Macrophage (S00V49N1)
Inflammatory Macrophage
Venous Blood Inflammatory Macrophage (S00H6O51)
Venous Blood Inflammatory Macrophage (S0022253)
Venous Blood Inflammatory Macrophage (S0015753)
Venous Blood Inflammatory Macrophage (S0019U51)
Alternatively Activated Macrophage
Venous Blood Alternatively Activated Macrophage (S00FTN51)
Venous Blood Alternatively Activated Macrophage (S00BS451)
Venous Blood Alternatively Activated Macrophage (S006VI53)
Venous Blood Alternatively Activated Macrophage (S0062252)
Inflammatory Macrophage (Cord Blood)
Cord Blood Inflammtory Macrophage (S007SK51)
Cord Blood Inflammatory Macrophage (S0018A52)
Cord Blood Inflammatory Macrophage (S00DVR51)
Cord Blood Inflammatory Macrophage (S00BHQ51)
Alternatively Activated Macrophage (Cord Blood)
Cord Blood Alternatively Activated Macrophage (S00E8W51)
Cord Blood Alternatively Activated Macrophage (S00C1H51)
Neutrophils
Venous Blood Mature Neutrophil (S00TC1A1)
Venous Blood Mature Neutrophil (S00K88A1)
Venous Blood Mature Neutrophil (S00K7AA1)
Venous Blood Mature Neutrophil (S00KSEA1)
Venous Blood Mature Neutrophil (S00FXFA1)
Venous Blood Mature Neutrophil (S00FW9A1)
Venous Blood Mature Neutrophil (C0011IA3bs)
Venous Blood Mature Neutrophil (C0010KA1bs)
Venous Blood Mature Neutrophil (C001UYA1bs)
Venous Blood Mature Neutrophil (C000S5A2bs)
Neutrophils (Cord Blood)
Cord Blood Mature Neutrophil (C0018AJbs)
Cord Blood Mature Neutrophil (C004OD51)
Band Form Neutrophils
Bone Marrow Band Form Neutrophils (S00VEQA1)
Bone Marrow Band Form Neutrophils (S00JMLA1)
Bone Marrow Band Form Neutrophils (S00JGXA1)
Segmented Neutrophils
Bone Marrow Segmented Neutrophils (S00VFOA1)
Bone Marrow Segmented Neutrophils (S00JNJA1)
Bone Marrow Segmented Neutrophils (S00JHVA1)
Monocyte
Venous Blood Monocyte (S01MAPA1)
Venous Blood Monocyte (S01E03A1)
Venous Blood Monocyte (S01BH0A1)
Venous Blood Monocyte (S01BGKA1)
Venous Blood Monocyte (S01BFMA1)
Venous Blood Monocyte (S01BEOA1)
Venous Blood Monocyte (S01BDQA1)
Venous Blood Monocyte (S01BCSA1)
Venous Blood Monocyte (S01BBUA1)
Venous Blood Monocyte (S01BAWA1)
Venous Blood Monocyte (S01B71A1)
Venous Blood Monocyte (S01B63A1)
Venous Blood Monocyte (S01B55A1)
Venous Blood Monocyte (S01B47A1)
Venous Blood Monocyte (S01B3YA1)
Venous Blood Monocyte (S00V2DN1)
CD14+, CD16- Classical Monocytes
Venous Blood CD14+ CD16- Classical Monocyte (C0010KA2bs)
Venous Blood CD14+ CD16- Classical Monocyte (C004SQS1)
Venous Blood CD14+ CD16- Classical Monocyte (C001UYA3bs)
Venous Blood CD14+ CD16- Classical Monocyte (C000S5A1bs)
Others
Eosinophils
Venous Blood Mature Eosinophils (S00V6S53)
Venous Blood Mature Eosinophils (S006KE53)
Dendritic Cells
Cord Blood Conventional Dendritic Cells (S00DT151)
Cord Blood Conventional Dendritic Cells (S00CP651)
Immature Dendritic Cells
Venous Blood Immature Conventional Dendritic Cells (S00TWZA1)
Venous Blood Immature Conventional Dendritic Cells (S00TV0A1)
Venous Blood Immature Conventional Dendritic Cells (S00TU2A1)
Mature Dendritic Cells
Venous Blood Mature Conventional Dendritic Cells (S00U0LA1)
Venous Blood Mature Conventional Dendritic Cells (S00TXXA1)
Venous Blood Mature Conventional Dendritic Cells (S00TYVA1)
Megakaryocyte
Cord Blood CD34+, CD41+ Megakaryocyte (S004BT51)
Cord Blood CD34+, CD41+ Megakaryocyte (S004AV51)
Erythroblast
Cord Blood Erythroblast (S0025351)
Cord Blood Erythroblast (S002R551)
Neutrophilic Myelocyte
Bone Marrow Neutrophilic Myelocyte (S00VCUA1)
Bone Marrow Neutrophilic Myelocyte (S00JJRA1)
Bone Marrow Neutrophilic Myelocyte (S00JE0A1)
Neutrophilic Metamyelocyte
Bone Marrow Neutrophilic Metamyelocyte (S00VD5A1)
Bone Marrow Neutrophilic Metamyelocyte (S00JKPA1)
Bone Marrow Neutrophilic Metamyelocyte (S00JFZA1)

Supplementary Figure 7 (**a**) Multiscaled view of Solo-WCGW methylation in iPSC and ESC-derived cells, showing deep PMD in H1-derived MSCs and residual PMD in iPSCs. (**b**) Multiscale view of Solo-WCGW CpG methylation in T, B and plasma cells of different varieties, showing deep PMD hypomethylation in regulatory T cells, germinal center B cells, memory T, B cells and plasma cells. (**c**) Multiscale view of Solo-WCGW methylation in myeloid cells, showing deeper PMD in megakaryocytes and erythroblasts.



271

272

273 **Supplementary Figure 8** Distribution of cross-sample SDs for solo-WCGW methylation

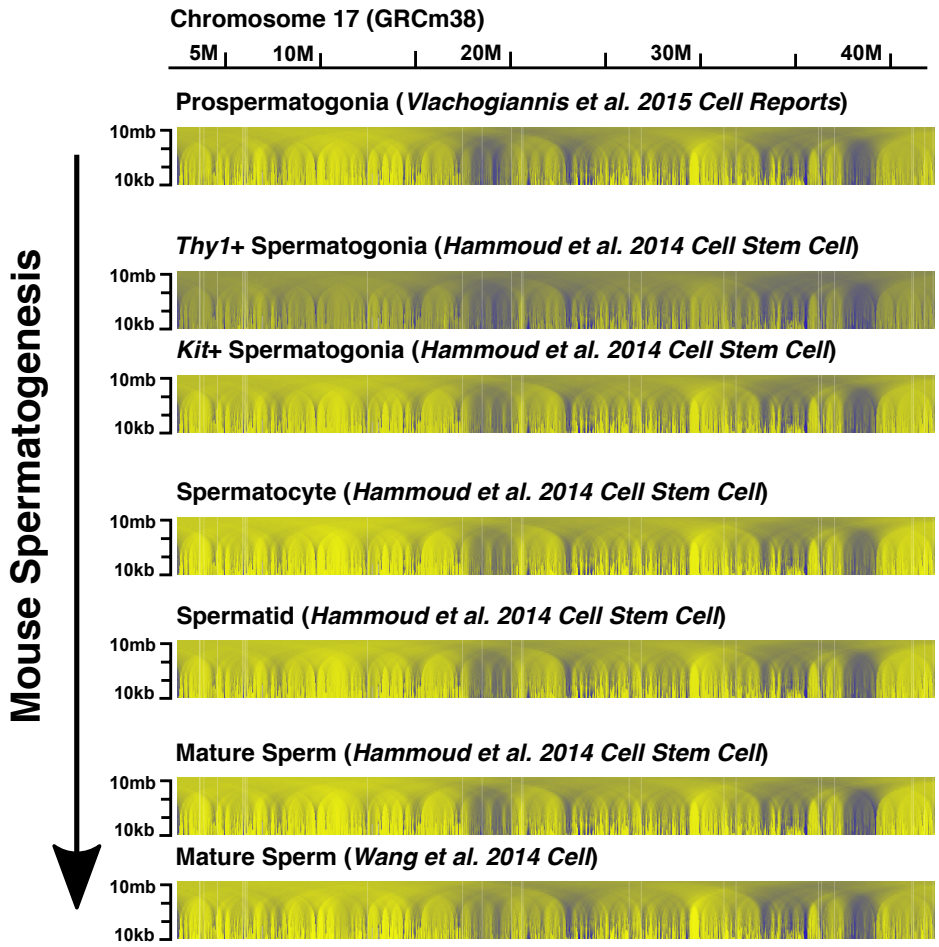
274 in all genomic 100kb bins of the core tumor group (studied in Fig 2b-c) plotted on Y-axis,

275 against SD distribution from (**a**) 50 other blood malignancies; and (**b**) 10 other solid

276 tumors, plotted on X-axis. The figure shows the concordance of SD-based PMD

277 definitions based on the core tumors and other tumors.

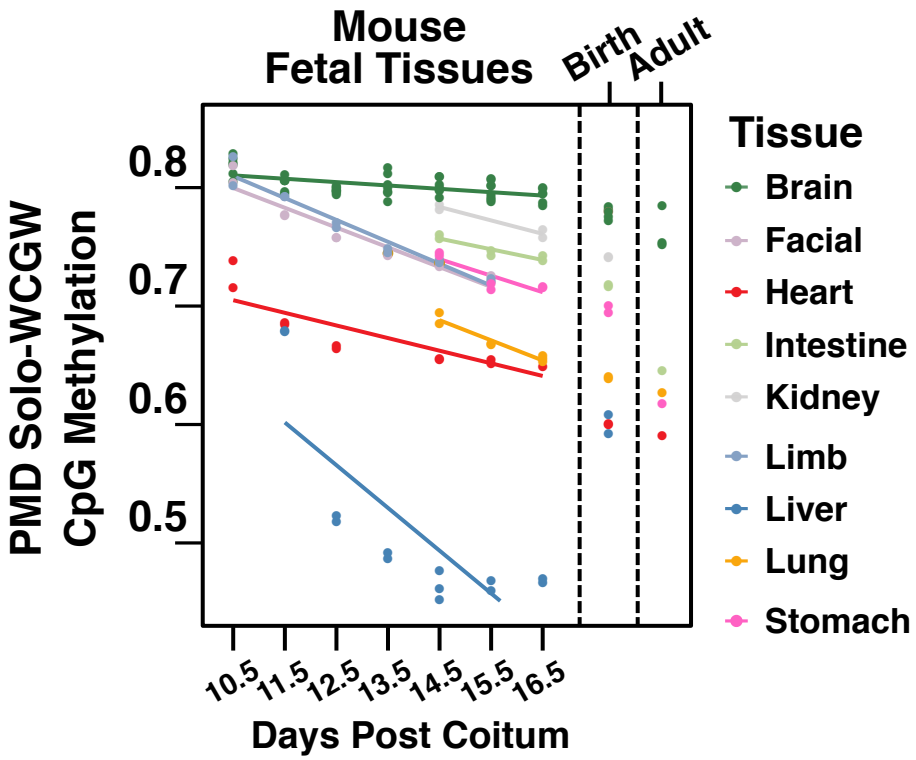
278



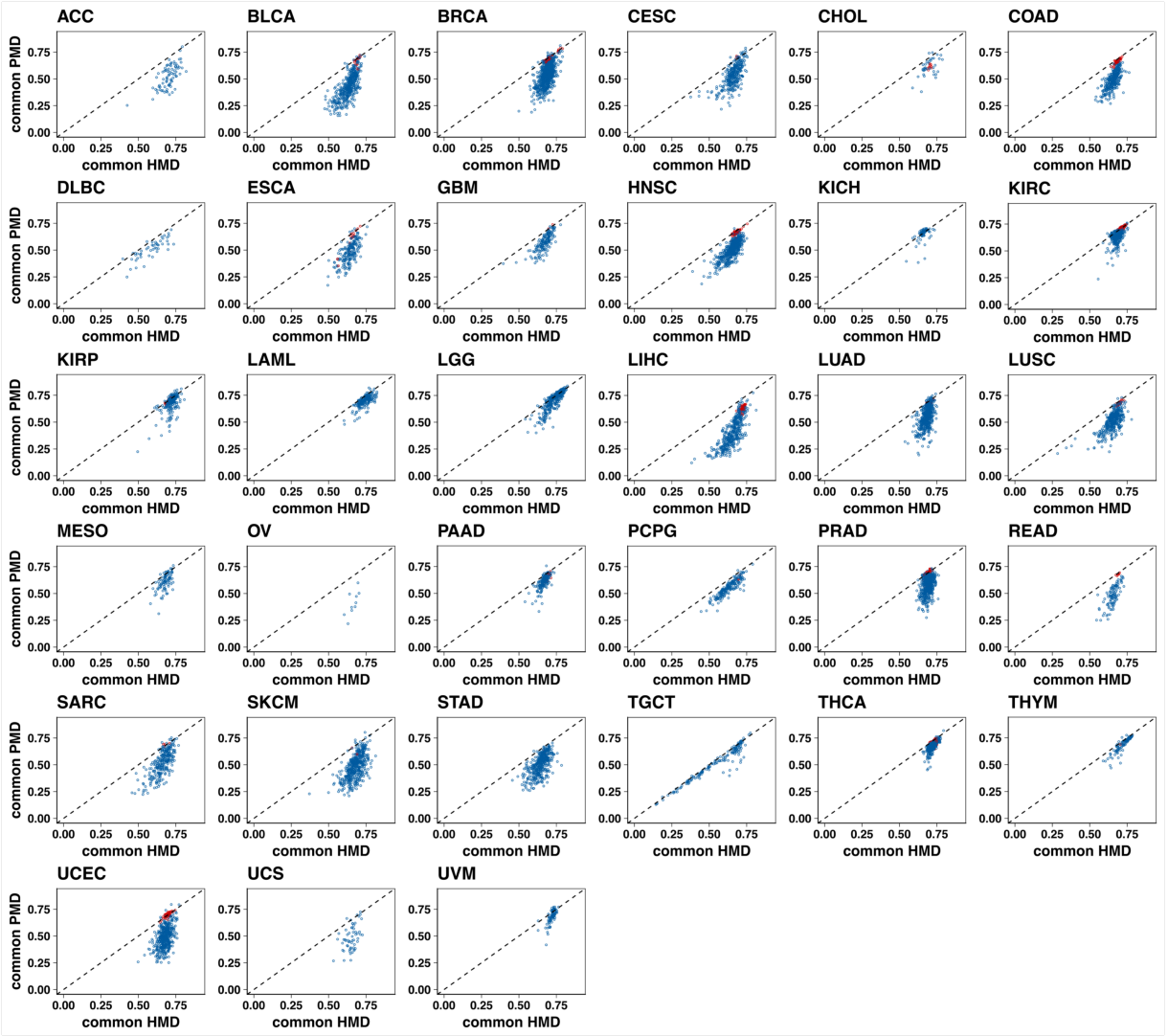
279

280 **Supplementary Figure 9** Multiscaled view of chromosome 17 (3-43Mbp) Solo-WCGW
281 methylation in different stages of mouse spermatogenesis from prospermatogonia to
282 mature sperm.

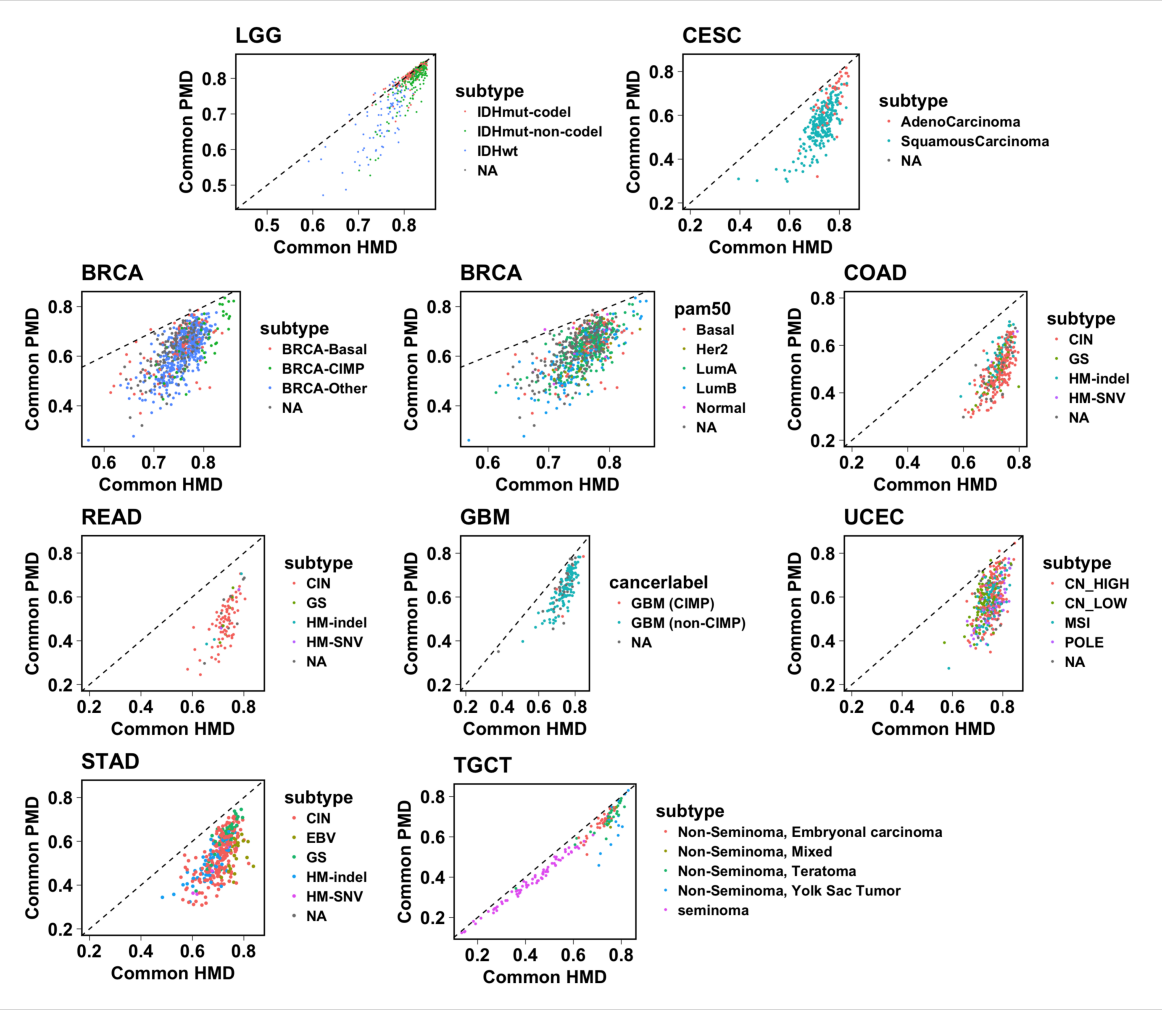
283



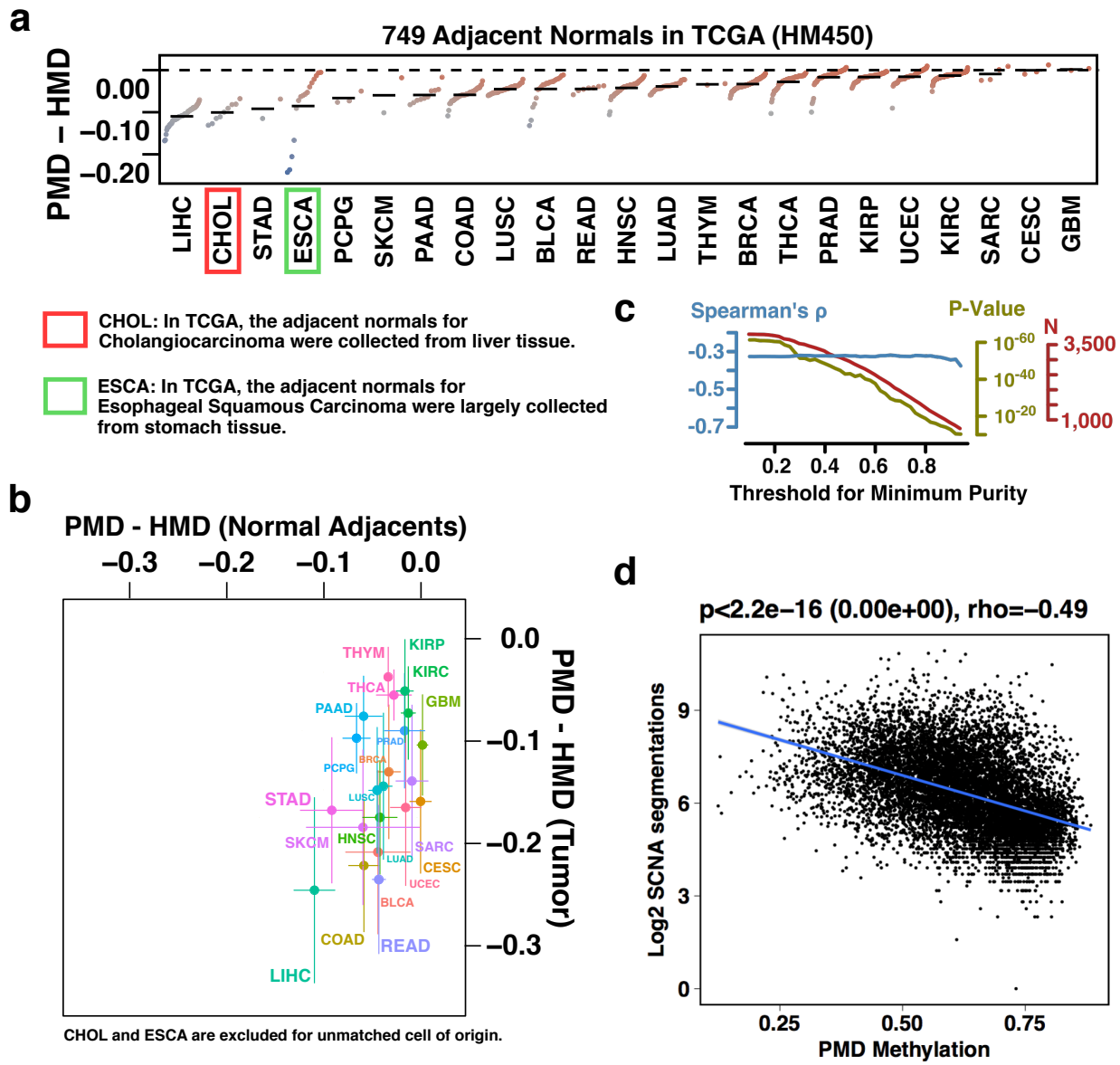
Supplementary Figure 10 Association of average PMD solo-WCGW CpG methylation with gestational age in mouse WGBS data sets stratified by tissue types.



Supplementary Figure 11 Solo-WCGW methylation average in common HMD and common PMD in 9,072 TCGA tumor samples from 33 tumor types.

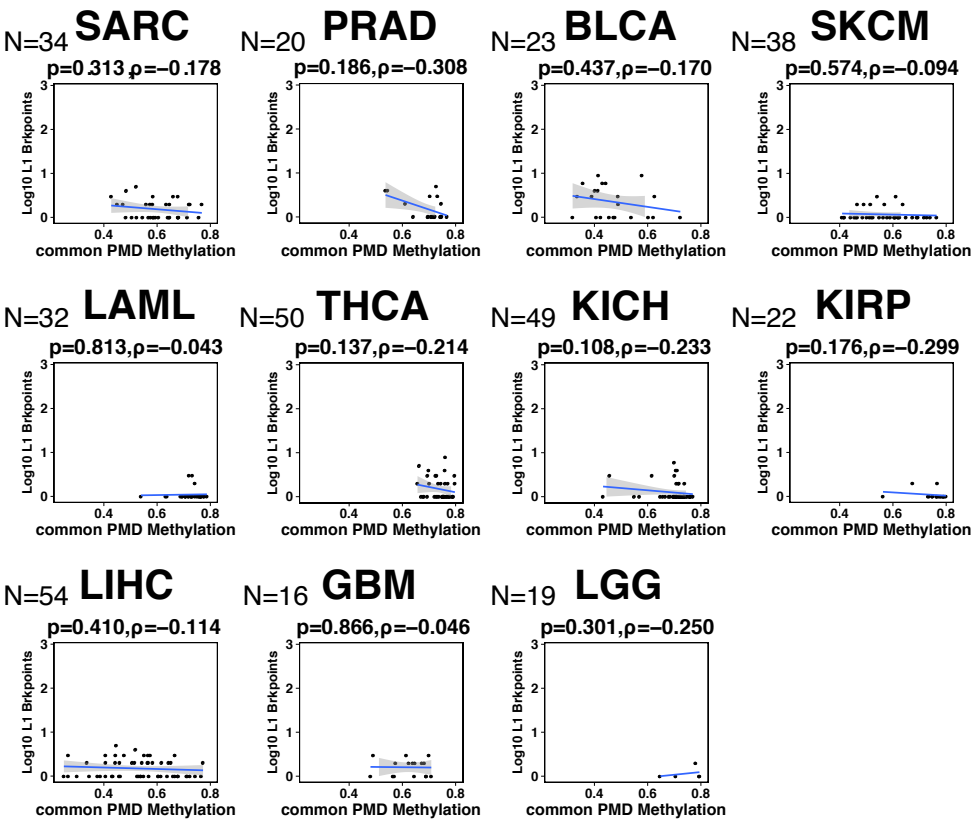


Supplementary Figure 12 Subtype-stratification of Solo-WCGW methylation average in common HMD and common PMD in TCGA tumor samples from 10 cancer types.

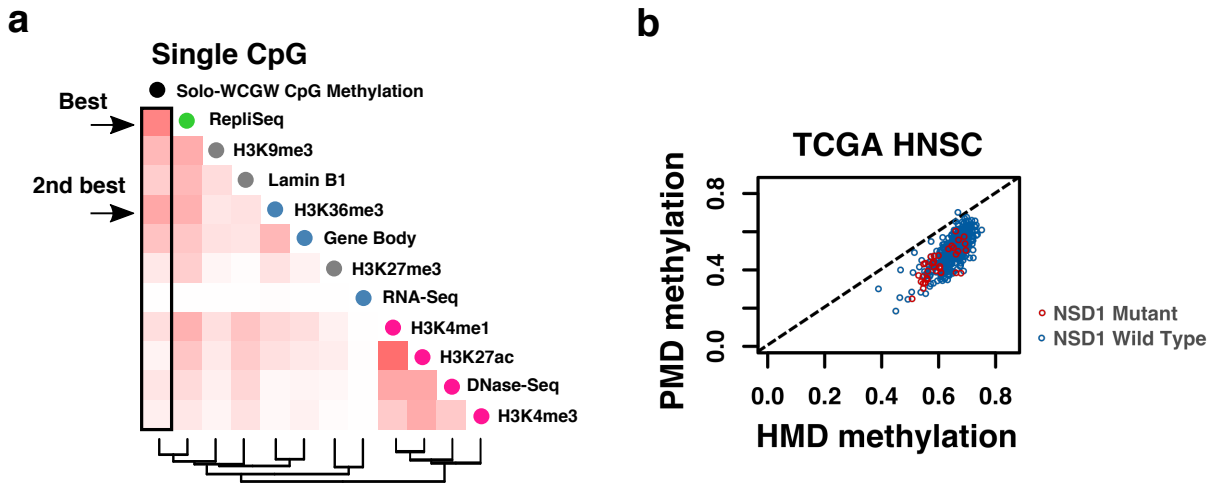


Supplementary Figure 13 (**a**) Difference of PMD and HMD methylation average of 6,214 Solo-WCGW probes in 749 adjacent normal samples assayed in TCGA on HM450 platform. (**b**) Comparison of normal (N=749) vs tumor (N=9,072) HMD-PMD methylation based on Solo-WCGW CpGs in 33 cancer types in TCGA with lines indicate standard deviation. The sample sizes are: ACC(N=80); BLCA(N=419); BRCA(N=799); CESC(N=309); CHOL(N=36); COAD(N=316); DLBC(N=48); ESCA(N=186);

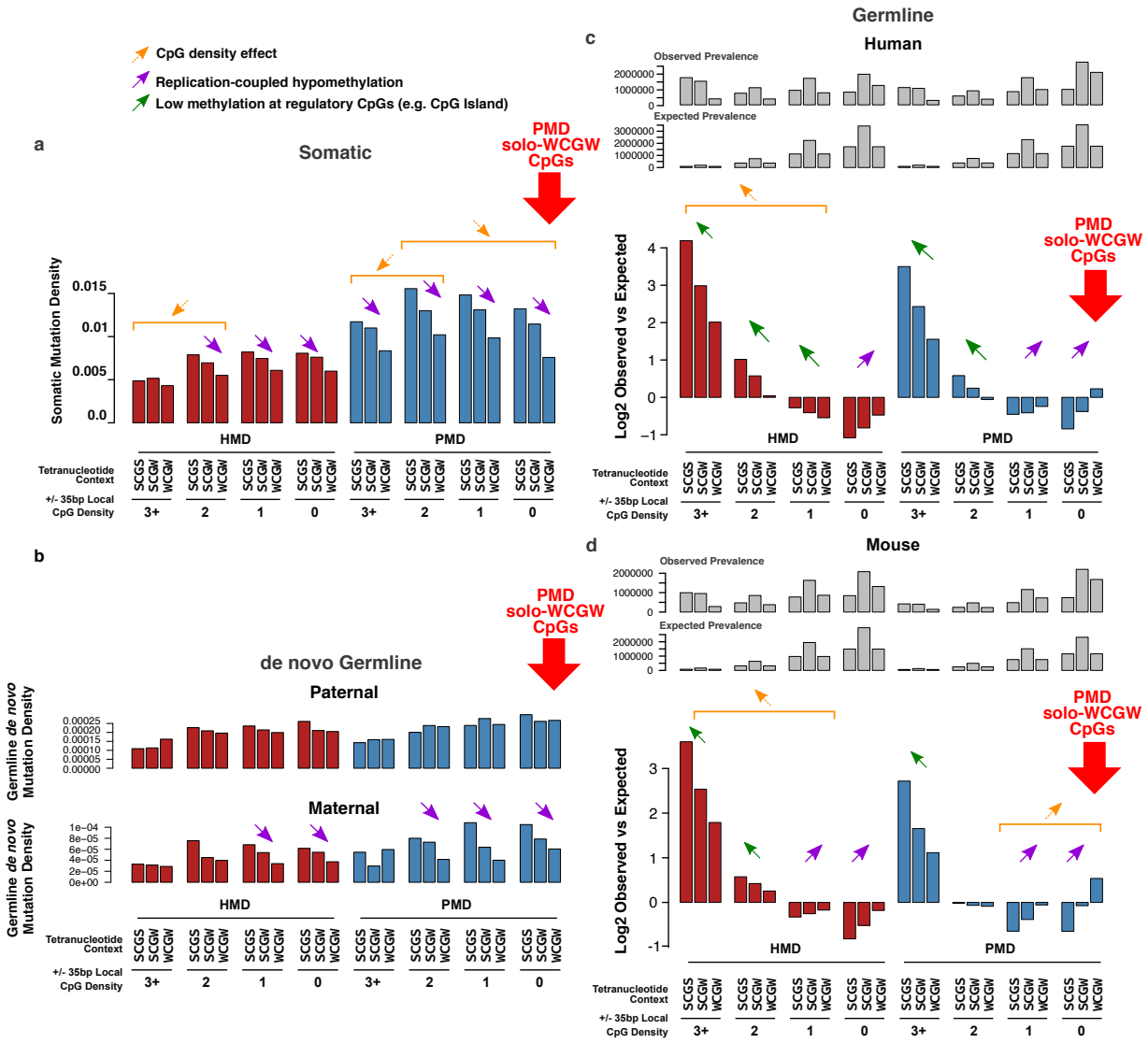
GBM(N=153); HNSC(N=530); KICH(N=66); KIRC(N=325); KIRP(N=276); LAML(N=194); LGG(N=534); LIHC(N=380); LUAD(N=475); LUSC(N=372); MESO(N=87); OV(N=10); PAAD(N=185); PCPG(N=184); PRAD(N=503); READ(N=99); SARC(N=265); SKCM(N=474); STAD(N=396); TGCT(N=156); THCA(N=515); THYM(N=124); UCEC(N=439); UCS(N=57); UVM(N=80); The sample sizes for normals are: BLCA(N=21); BRCA(N=98); CESC(N=3); CHOL(N=9); COAD(N=38); ESCA(N=16); GBM(N=2); HNSC(N=50); KIRC(N=160); KIRP(N=45); LIHC(N=50); LUAD(N=32); LUSC(N=43); PAAD(N=10); PCPG(N=3); PRAD(N=50); READ(N=7); SARC(N=4); SKCM(N=2); STAD(N=2); THCA(N=56); THYM(N=2); UCEC(N=46); The mean of each data set is used to measure the center; (**c**) Spearman's correlation coefficient for analysis in (**Fig. 7b**), shown as a function of minimum purity threshold from 0.1 to 0.95 (hypermutators excluded, Online Methods). PMD hypomethylation in TCGA tumors was captured by the average DNA methylation beta values of common PMD HM450 probes. (**d**) Correlation between PMD methylation (average DNA methylation beta value of HM450 common PMD probes) and the number of Somatic Copy Number Aberration (SCNA) in TCGA tumor sample (N=9454).



Supplementary Figure 14 Association of LINE-1 break points and PMD methylation (characterized by average of HM450 probes in common PMDs). Rho is Spearman's correlation coefficient. P-value was calculated using algorithm AS89 implemented in the R software.



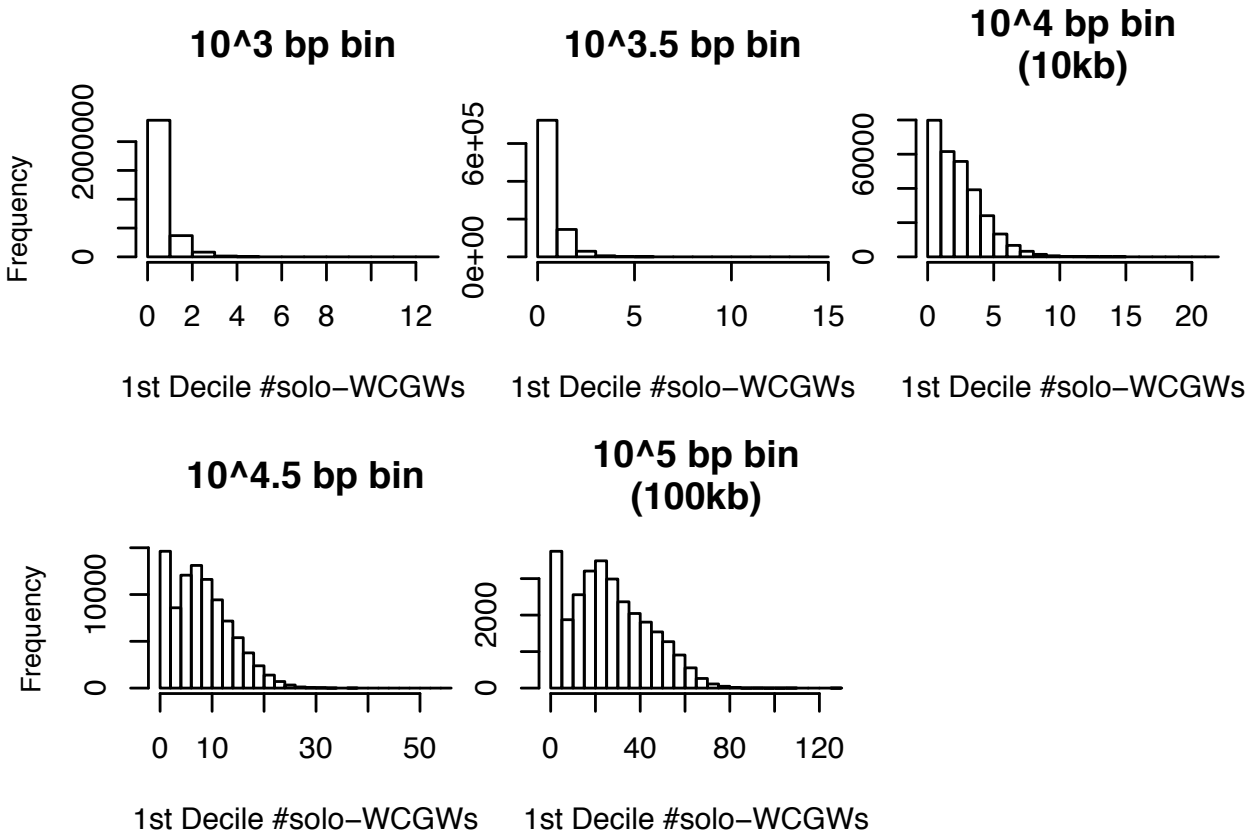
Supplementary Figure 15 (**a**) Spearman correlation coefficients of Solo-WCGW CpG methylation and 10 other epigenomic features of IMR90 fibroblast at single CpG scale; Samples were hierarchically clustered based on distances defined by 1- abs(rho). The dendrogram of clustering is shown on the bottom with arrow indicating the best and the 2nd best correlator with Solo-WCGW CpG. (**b**) PMD vs HMD methylation average of Solo-WCGW HM450 probes in TCGA HNSC tumors showing NSD1 wild types and mutants.



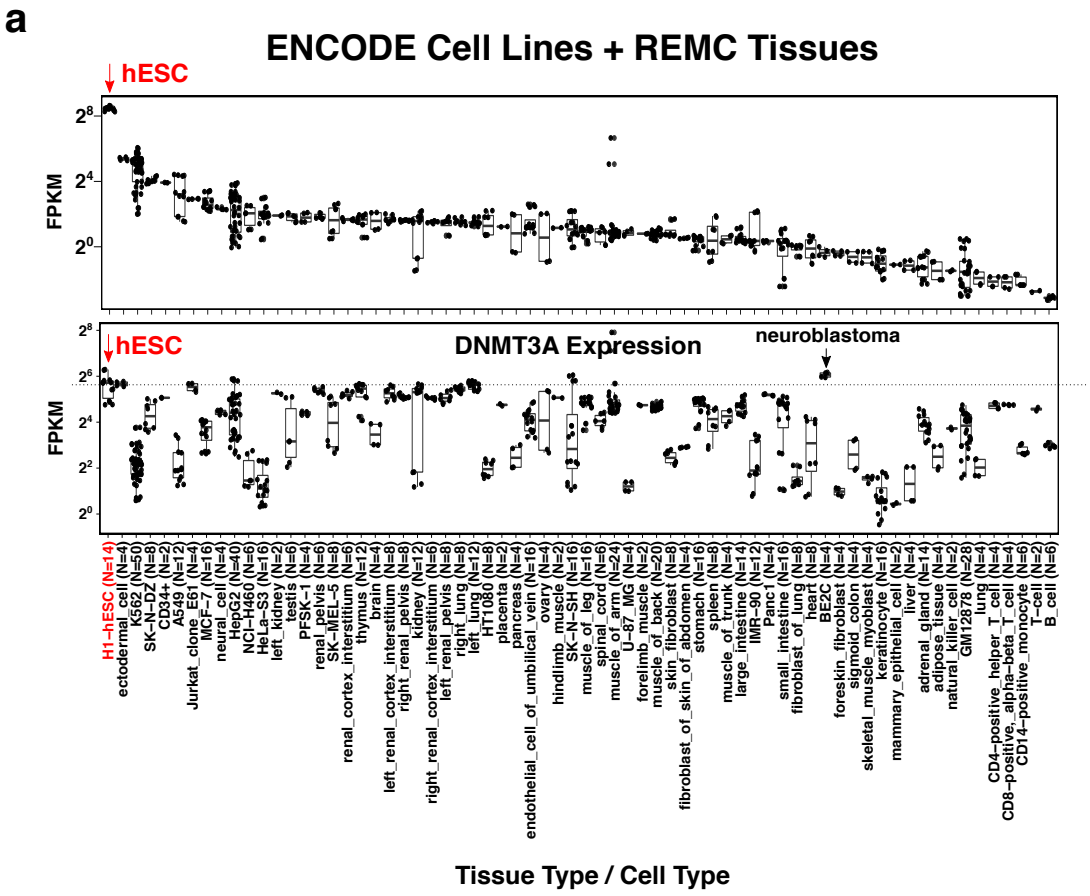
344

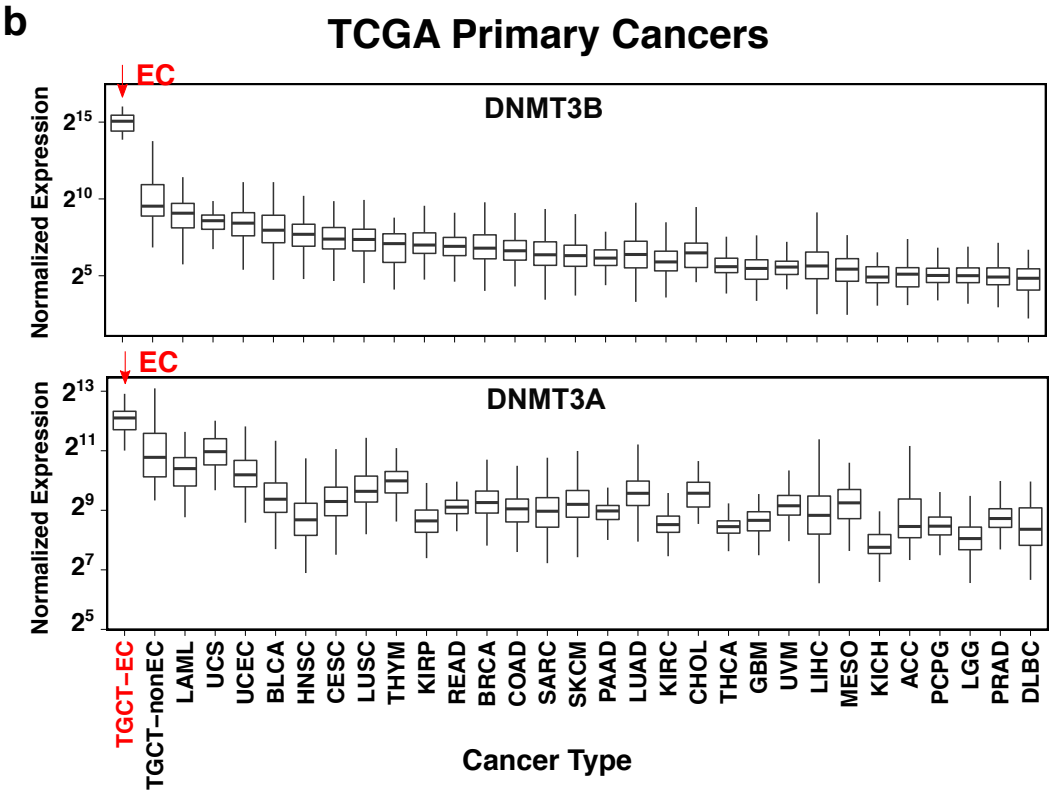
345 **Supplementary Figure 16** (**a**) Impact of CpG dinucleotide PMD/HMD location, flanking
346 CpG density and tetranucleotide sequence context on somatic mutation rate in 100
347 gastric cancer WGS[24]; (**b**) Impact of CpG dinucleotide sequence context on *de novo*
348 germline mutation rates estimated from 1,548 Icelandic trios[25]; (**c**) Genomic CpG
349 distribution stratified by PMD/HMD, flanking CpG density and sequence context in
350 human; (**d**) Genomic CpG distribution stratified by PMD/HMD, flanking CpG density and
351 sequence context in mouse.

352

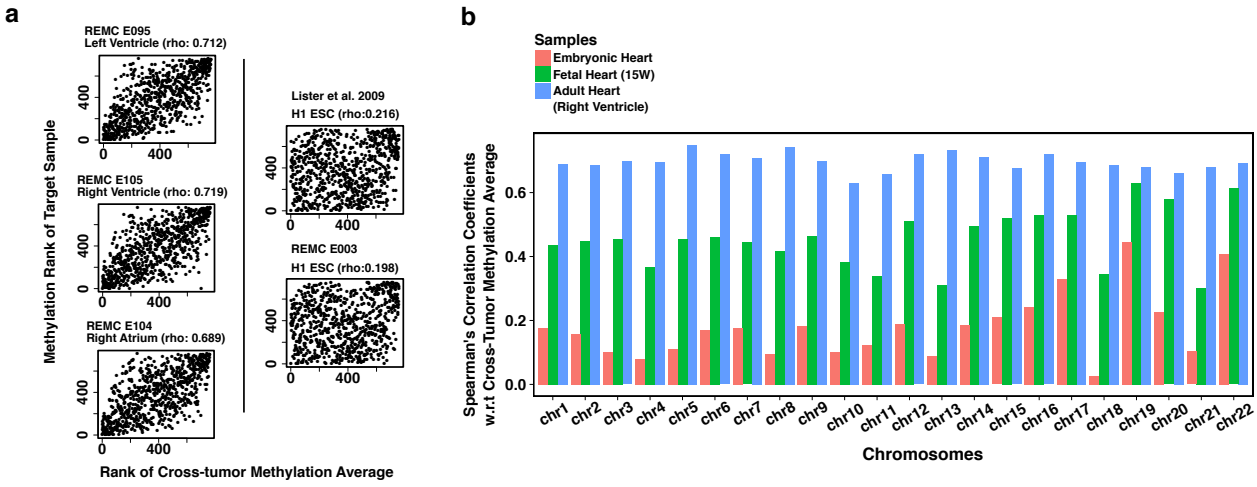


Supplementary Figure 17 First decile of the number of solo-WCGW CpGs in windows of different sizes that were used to segment the whole genome;


a
ENCODE Cell Lines + REMC Tissues
hESC
FPKM
DNMT3A Expression
neuroblastoma
Tissue Type / Cell Type


b
TCGA Primary Cancers
EC
Normalized Expression
DNMT3B
DNMT3A
TGCT-EC
TGCT-nonEC
LAML
UCS
UCEC
BLCA
HNSC
CESC
LUSC
THYM
KIRP
READ
BRCA
COAD
SARC
SKCM
PAAD
LUAD
KIRC
CHOL
THCA
GBM
UVM
LIHC
MESO
KICH
ACC
PCPG
LGG
PRAD
DLBC
Cancer Type

Supplementary Figure 18 mRNA expression of DNMT3A and DNMT3B in (**a**) ENCODE cell lines and Roadmap Epigenome Consortium (REMC) primary tissues (each data point corresponds to the expression level for a cell line or primary tissue type) and (**b**) All TCGA cancer types with TGCT split into tumors of the embryonic origin (TGCT-EC) and non-embryonic origin (TGCT-nonEC). This figures shows elevated DNMT3B expression in hESCs and embryonic carcinomas compared to other tissues and cancers by over an order of magnitude. Each data point in the box plot represents the normalized expression level for a cancer sample. Samples sizes for all cancer types are: ACC(N=79); BLCA(N=427); BRCA(N=1218); CESC(N=310); CHOL(N=45); COAD(N=329); DLBC(N=48); GBM(N=174); HNSC(N=566); KICH(N=91); KIRC(N=606); KIRP(N=101); LAML(N=173); LGG(N=534); LIHC(N=424); LUAD(N=576); LUSC(N=554); MESO(N=87); OV(N=266); PAAD(N=183); PCPG(N=187); PRAD(N=550); READ(N=105); SARC(N=265); SKCM(N=473); TGCT(N=156); THCA(N=572); THYM(N=122); UCEC(N=201); UCS(N=57); UVM(N=80);



Supplementary Figure 19 (**a**) Rank correlation between three closely-related heart tissues and two replica of H1 ESC from different studies showing the magnitude of variation; N=792 non-overlapping 100kbp genomic windows in chromosome 16. (**b**) Order of Spearman's correlation in different chromosomes between the core tumor samples and the heart tissue samples from three different developmental stages.